Molecular genetic strategies for species identification

XU Yan-chun (徐艳春)*, BAI Su-ying (白素英), JIN Yu (金煜), JING Song-yan (景松岩)
(College of wildlife Resources, Northeast Forestry University, Harbin 150040, P. R. China)

Abstract: This paper probes into the molecular genetic mechanism of the formation of species, subspecies and variety in evolving progression, and brings forward 5 criteria of an ideal strategy in species identification: stating the specific characteristics at species, subspecies and variety level without any interference of too high polymorphism at individual or population level; keys should be distributed as 0 or 1, e. g. yes or no; satisfying repeatability and simple operation; high veracity and reliability; adaptability to widely various specimen. Respectively, this paper reviews two strategies focusing on detecting the fragment length polymorphism and base replacement and lays out some detail methods under above strategies. It demonstrates that it is not possible to solve all species problems by pursuing identification with only a single gene or DNA fragment. Only based on thorough consideration of all strategies, a method or combined several methods could bring satisfying reliability. For advanced focuses, it requires not only development and optimization of methods under above strategies, but also new originality of creative strategies.

Key word: Species identification; Molecular genetics; Forensic science
CLC number: Q75 **Document code:** A **Article ID:** 1007-662X(2000)04-0249-03

Introduction

Wildlife management has been brought forward and upgraded by all nations. Multilateral agreements like 'Conventions on International Trade of Endangered Species of Wild Fauna and Flora (CITES)' enhanced investigation of international contraband of wildlife. 'Law of Wildlife Protection' and local regulations provide a legislative base for cracking down illegal poaching, trading and transporting wildlife. However, the profit of wildlife contraband stands the third high after the contraband of ammunition and drug. A fanum is made of Tibetan antelope fuzzy hair could be sold for 40 thousand dollars. Specimen of other world grade endangered species like tiger, golden monkey, and giant panda could be even higher. This benefit leads to a high strength of poaching, and poaching makes species more endangered, the more endangered or rare the species is, the more benefit can be gained in poaching. There is a hard challenge to stop the bad circle. In such cases, involved species, subspecies and varieties must be identified from living body, art crafts, traditional Chinese medicine, fur or hair products, leather products, unprocessed animal parts etc. and determined if they are national or international controlled. With the development of stockbreeding, the same identification is involved in more and more criminal and civil cases. The key to identify species is to find out the interspecies differences that are significant and stable enough. It attracted the attention of biologists and forensic scientists. Since 1970s, molecular genetics and relative technologies cut new ways to answer these questions.

Interspecies difference originates from difference of DNA sequence

In the evolving process, DNA base mutation provides chances for species differentiation. When base mutation runs up to a critical degree to form a reproductive isolation between two populations, two new species are thus formed. In any population, mutation always takes place on DNA at certain frequencies. One of four base kinds can be replaced by other three kinds. Such replacement will lead to changes of some amino acid codons and change the primary structure of protein molecules when it happens in coding regions. If changed protein molecule remains the original secondary structure and functions, the mutation will be acquired by the genome and maintained in next generations. Sometimes, the changes in primary structure cause changes in secondary structure and functions. If the individuals with new functions of mutated proteins show better adaptability in some aspects, such mutations will be maintained by more and more individuals through natural elimination of those who do not mutate, especially in cases of natural disasters. Otherwise, individuals with the mutations that cannot benefit them or hurt

Biography: *Xu Yan-chun (1970-), male, Ph. Doctor, lecture in Northeast Forestry University, Harbin 150040, P. R. China.
Received date: 2000-05-25
Responsible editor: Song Funan

their adaptability will be eliminated under the selective pressure unless other alleles can reinforce the functions. As a result, such mutations are hard to be maintained amongst a population. Selection and accumulation of mutations finally divaricated populations that lie in different selective system in genetics. When the divergence reaches a certain level enough to isolate reproduction, two new species are thought formed. Genetic divergence is detectable at protein level. But protein only changes after DNA quit a little conservatively, so DNA detection reflects more information. That is why DNA came to be the most favored to detect genetic divergence.

Base replacement also, more frequently, takes place in noncoding regions where protein structures are not relevant, seeming silent and ineffective. Natural selection does not squint towards one mutation prior to others unless it is linked with encoding regions. Genome accumulates much more base replacements in noncoding regions and introns than coding regions and extrons.

Besides base replacement, fragment insertion and deletion are other two types of mutation. Insertion and deletion are seen more frequently in noncoding regions and introns than coding regions and extrons because of the same reason as base replacement that mutations relevant to the protein structure and functions are strictly selected. As a result of insertion and deletion, fragment length can be lengthened and shortened.

Mutations as above appear on individual genome as individual polymorphisms. By comparing different populations, some mutations are to be found prior to others in one population--gene frequency and genotypes show population specialty. Such specialities are seen more significant as different subspecies, species and upper taxonomic orders are compared.

Detecting strategies for species divergence

Macroscopically, there are many different aspects among species in genotypes like morphology, physiology and behavior etc. Difference in DNA that must be more significant provides extensive space to search for the specialities representing a unique species, subspecies and varieties. Reported researches demonstrate that DNAs encoding structural proteins do not show speciality because any animal must live on common basic structure, like muscle, hair, skin, claw, eyes and ears etc. In contrast, DNAs encoding functional proteins and not encoding proteins do show species specific specialities. Detecting strategies should focus on later DNAs.

An ideal method for species identification should accord with the following criteria:

a) Stating the specific characteristics at species, subspecies and variety level without any interference of too high polymorphism at individual or population level;
b) Keys should be distributed as 0 or 1, e. g. Yes or No;
c) Satisfying repeatability and simple operation;
d) High veracity and reliability;
e) Adaptability to widely various specimens.

For the first 2 criteria, design must be based on thorough survey of all populations belonging to one species. The later 3 ones can be settled by optimizing detail scheme. All schemes focus on polymorphic loci in genome, such as satellite, minisatellite, microsatellite etc. But some loci, like microsatellite contain too high polymorphisms that distribution of allele frequency of two species overlaps and the probability of exclusion and inclusion moves downward when the overlapped area increases. Therefore, strategies must make for sure that the first 2 criteria are fulfilled.

Strategies and methods of species identification in common use

PCR based detection of fragment length polymorphisms

In PCR method, primers are designed flanking target DNA to duplicate millions times of its copies. Variable region of 28s rRNA gene (Naito *et al.* 1992) and 3' untranslated region of *SON* gene (Soteriou *et al.* 1995) are typical species specific loci. Conservative regions common to any species flanking the variable region in each locus are taken as primers. The length polymorphisms of PCR products demonstrate species.

Another countermeasure to design primers is to make the primers only anneal with DNA of specific species and yield products. Wetton *et al.* (1999) employed this strategy to determine tiger bone in Chinese patent medicines. They firstly sequenced and compared a part of cytochrome b gene of mtDNA in many felids. In tiger DNA, a region of which DNA sequence is similar with other cats except for 1 or 2 bases were taken as PCR primers. Once there really exists tiger DNA in the DNA mixture isolated from Chinese patent medicines, the primers always give products, while give nothing in other species because of difference of 1 or 2 bases between primers and corresponding regions. The accumulation of species specific amplification products were monitored in real-time by the increased fluorescence of a dye that only emits lights when bound to double stranded DNA. A single peak in monitor graphs means a tiger component. If a common primer pair was added into the reaction system as positive control that amplifies successfully under the same reac-

tion conditions, common electrophoresis of polyacrylamide or agarose gel can be used to detect PCR products. By this means, this strategy can be extended widely. But its reliability must be based on thorough survey in mtDNA sequences in a large number of species.

PCR based detection of base replacement

Species differentiation is exhibited as the diversity in DNA sequence. A DNA fragment not showing fragment length polymorphism among species may contain large divergence in base sequence. mtDNA regions like cytochrome b gene, amino acid tRNA gene and control region, nuclear DNA regions like introns and noncoding regions are typical regions accumulating base replacement. Detecting base replacement by targeting these regions can meet the demand of species identification.

DNA sequencing technique is the most effective means to detect base replacement. But sequencing shows too high polymorphism at individual level that conceals to species. So we must think over other means only reveal the specific keys. A valuable approach is to detect if base replacement creates or removes digest sites using restriction enzymes. Base replacements are not detectable if no digest site of restriction enzyme is created or removed. Restriction fragment length produced by this means is monomorphic within a species and polymorphic among species. This approach effectively reduces the impact of too high individual polymorphisms and obtains species specific keys.

US Fish and Wildlife Forensic Laboratory (1997) has amplified cytochrome b gene of mtDNA in American black bear (*Ursus americanus*), Brown bear (*Ursus arctos*), polar bear (*Ursus maritimus*), sun bear (*Helarctos malayanus*) and Asiatic black bear (*Ursus thibetanus*) using primer L14724 and H15149. PCR products were cut with *Rsa I* and *Alu I*. AFLPs banding pattern on polyacrylamide gel showed difference among species. Cutting the PCR products of the same primer pair with *Rsa I*, they disclosed the difference among white tailed deer (*Odocoileus virginianus*), mule deer (*Odocoileus hemionus*), red deer (*Cervus elaphus*) and rein deer (*Rangifer tarandus*).

Presently, methods and strategies mentioned above have some defects. In the detection of fragment length polymorphisms, fragment length variation sometimes is found very tiny. The length of variable region of 28s rRNA gene is about 100bp, while the length difference between rabbit and chick is only 2bp. Common electrophoresis techniques can not detect such difference with a high reliability. In addition, 28s rRNA gene is a multiple copied gene among genome and always producing multiple products. Without reliable quality control, it's hard to determine whether a band is a target band or not. The length of 3' untranslated region of *SON* gene is 175bp in human, 174bp in dog and 172bp in domestic cat respectively. It has too much adventure to isolate such tiny difference by common electrophoresis being used in most labs. Moreover, fragment length in some species is exactly the same. In the strategy amplifying DNAs of specific species, it's very hard to tell the exclusion from experiment failure when negative results are acquired without positive control. Methods of AFLPs must be based on investigation of large scale of samples before it could be applied. So it is now only applied in ursid and cervid animals.

With the development of agricultural technology, disputes involving variety and subspecies of domestic animals and wildlife under artificial breeding appear more frequently. Reports relative to distinction of subspecies and variety include microsatellite polymorphism analysis, RFLPs, AFLPs and DNA sequencing. But these reports emphasize subspecies and variety differentiation more than providing keys to identification. The difference among subspecies and varieties are presented statistically. For one sample, the inclusive or exclusive probability is not higher than 60-70%. This is not acceptable in casework.

Up to now, it hasn't been found that a method that can be applied to solving all problems of identification of species, subspecies and varieties. From the point of evolution, it seems impossible to create such a method by only one gene or DNA fragment. We must integrate the advantages and disadvantages of all strategies and methods in the considerations and combine several ones in a deal. Henceforth, attentions should not only be paid to optimizing above strategies and methods, but also to creating new originalities.

References

Naito, E., Dewa, K., Ymanouchi, H., Kominami, R. 1992. Ribosomal ribonucleic acid (rRNA) gene typing for species identification [J]. Journal of Forensic Science, **37**(2): 396-403.

National Fish and Wildlife Forensic Laboratory. 1998 Wildlife genetics DNA methods guide. (unpublished).

Soteriou, B., Fisher, R., Khan, A.I.M., *et al.* 1995. Conserved gene sequences for species identification: PCR analysis of the 3' UTR of the SON gene distinguishes human and other mammalian DNAs [J]. Forensic Science international, **73**: 171-181.

Wetton, J., Chan, C., Roney, C., Spriggs, A. 1999 Developing a DNA test for the identification of tiger bone. Forensic Science Society. (unpublished).